

Bioinformatics approaches for drug repositioning

Nicolas Jaume^{1,2}

Master 2 research BBSG, Faculty of Science¹, Aix-Marseille University, 163 Avenue de Luminy, 13288 Marseille, France

I2M, Marseille Institute of Mathematics² UMR 7373, 163 Avenue de Luminy, Case 901, 13288 Marseille, France

Abstract

Drug repositioning offers many advantages by reducing risks in drug discovery, saving time and money. It accounts for approximately 30% of the newly US Food and drugs administration FDA approved drugs in recent years. The rapid development of high-throughput technologies that can now be used for drug repurposing strategies implicates bioinformaticians whom have to develop innovative computational methods into their pipelines. In this review, we describe bioinformatics approaches applied to the process of drug repurposing at three different levels targets, -omics, disease and side effects.

Introduction

The drug discovery process is divided in three stages: i) discovery process where new compound are screened, ii) pre-clinical development where compounds are tested *in vitro* and in animal models and iii) clinical trial where drug candidates are tested on humans. The discovery process involves lead identification followed by chemical optimization. Pre-clinical studies analyze pharmacodynamics, pharmacokinetics, ADME and toxicology parameters and interaction with other drugs for the first-in-man study. Finally, clinical trials test the real drug effects on human beings generating the efficacy and safety data of the drug. Serious side-effects and decreased efficiency are the major cause of approval failure at this ultimate stage. Even if the number of drugs in Phase III clinical trials has doubled this past few years, new substances into development phases have been declining by 50% compared to 5 years ago. This productivity problem is coupled with a high concurrency context on prizes and challenges with generics drugs, which leads companies to find new drugs. In this context, novel discovery technologies are nowadays necessary.

The process of finding new therapeutic usages for existing drug, also known as drug repositioning or repurposing, is an attractive alternative for pharmaceutical and biotechnologies industry. Indeed, the drugs are already approved for treatment of some diseases or have validated pre-clinical studies. *De novo* drug discovery is a 10 to 17 years process from idea to marketed drug. It seems obvious that repositioning of known safety profiles

from an accelerated R&D process is less cost effective than conventional drug discovery and takes only 4 to 10

years (Figure 1). It allows bypassing the high risk of failure rate due to safety reason because of a proven bioavailability. In recent years, drug repositioning has represented 30% of the newly US Food and Drug Administration approved drugs¹⁸. The first big success of drug repositioning with sildenafil (VIAGRA®) from angina to erectile dysfunction raised high interest from pharmaceutical and biotech companies, resulting in the creation of startups exclusively focused on drug repositioning such as Pharnext²⁰ in France or biovista²¹ in the US. Following this success, nowadays, more than 150 drugs have been repositioned (Tab1) and the number of drug repositioning methods has severely increased.

Indeed, the development of high-throughput technologies nowadays gives a large amount of data and details on drug-disease mechanisms. This enables researchers, and in particular bioinformaticians, to develop efficient drug-repositioning approaches. The development of these approaches is tightly linked to the management of large-scale data, and to the integration and exploitation of information. In this review, we will describe bioinformatics approaches applied to the process of drug repurposing at three different levels. First, bioinformatics modeling strategies have been used to predict new drug usages based on the knowledge on their molecular targeting. Second, we will show how -omics data and associated bioinformatics tools can be integrated to discover drug repurposing strategies. Finally, we will show how the

knowledge at the phenotypic level can also offer a complementary approach to drug repositioning.

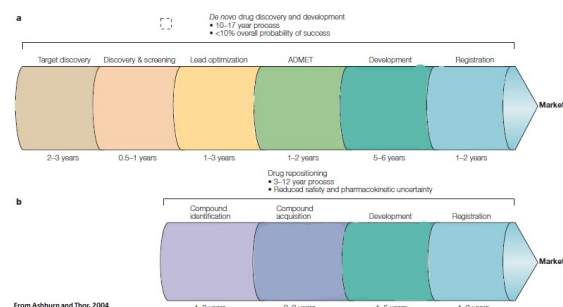


Figure 1. Comparison of de novo drug discovery and development versus drug repositioning a. De novo drug discovery process takes 10 to 17 years of development with less than 10% of probability of success. **b.** Drug repositioning process reduces time for development with proven safe compound.

Target-based approaches

When drug targets are already known because directly linked to the disease mechanisms, an *in silico* drug discovery process could be developed by undertaking virtual screenings thanks to bioinformatics tools and high computational power. Advantages of these target-based methods are that they enable researchers to screen all their chemical compounds libraries or drugs in a few days comparing to the classic experimental approach that could take several month or years. Companies like Genentech²², Melior discovery²³ and Aurora Biomed²⁴ are using these type of approaches.

Structure-based virtual screening method

Structure-based virtual screening or molecular docking is a high-throughput screening method. It allows the evaluation of the interactions between the targets and large libraries of compounds in a quick and cost-effective manner. In this context, drugs and compounds already approved can be systematically screened against new targets. Moreover, molecular docking tools can also be used to study the binding mode of small molecules to the target binding site. Several methods are used to work on ligand flexibility for building conformers before docking. The incremental building is based on geometric fitting, where ligands are partitioned into small fragments that are individually docked into the receptor pocket. This is followed by the incremental checking of the entire ligand. In the second method, several low-energy conformations of the same ligand are generated, followed by their individual docking against the receptor pocket⁶.

Molecular similarity method

The aim of the molecular similarity computational targeting is to identify the most likely targets of a query molecule and to get a prediction of the bioactivity and mechanism of action (MOA) of the query molecule. These approaches could permit to reposition drugs based on the “chemical similarity principle,” where similar molecules are likely to have similar properties. The advantage is that it only requires the computation of the similarity between compounds as described in Figure 2. The approach followed by Keiser *et al.*¹³ is first, a small molecule is represented as a chemical fingerprint that is a series of binary digits (bits) representing the structure of the molecule. Second, the Tanimoto coefficient is applied to compare the fingerprints of two molecules. The more similar two compounds are, the closer the Tanimoto coefficient will be to 1. A database containing the bioactivities of thousands compounds and drugs target properties are used to determine a potential target as described by Eming *et al.*¹⁵. However, inactive compounds can sometimes show good similarity with active molecules, this represent a limitation of this method. Third, it is important to determine the statistical significance of the similarity score of a query compound in the database¹⁵. Finally, two structures are considered as similar if the Tanimoto coefficient between them is higher than 0.85. The last step is to use the SwissTargetPrediction online tool to estimate the accuracy of the predictions²⁵.

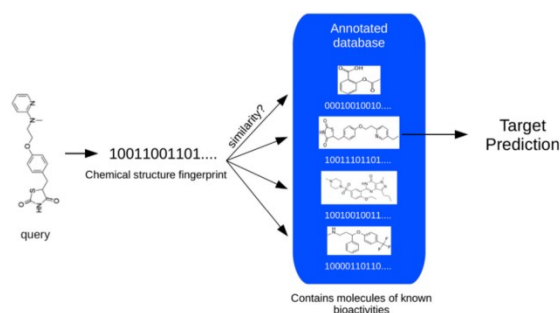


Figure 2. Chemical similarities through the chemical fingerprints to predict novel target. Fingerprints are a way of encoding the structure of molecule (drugs) in bits in order to compare their fingerprints to a database containing others drugs including target prediction properties. From Cereto-Massagué *et al.*, *Methods*, 2014

Predictive QSAR method

Predictive QSAR “Quantitative structure-activity relationship” method is based on the principle that similar

molecules have similar activities and small difference on molecular level lead to a different activity. Based on supervised machine learning approaches¹, the correlation between chemical structure particularities of compounds with their specific biological activity could predict new potential targets for bringing out novel drug–target interactions.

Omics based approaches

The link from genotype to phenotype is complex, such as the link from a molecule to the clinico-pathological state¹⁹. In recent years, many new technologies called -omics, which represent the “big data revolution” of biology, have been developed to decipher the links between genotypes to phenotypes, and to reflect environmental effects. Proof-of-Concept studies and clinical investigations have generated large amount of data on signaling, metabolism, interactome, protein-protein interactions related to diseases and healthy states. All these data are easily accessible from online databases such as GEO²⁶ (Gene expression omnibus) a big high-throughput functional genomic data sets or cMAP^{27,11} (connectivity Map) which is the largest public database of genome-wide expression profiles from five different human cancer cell lines treated with more than 1000 bioactive compounds. Nowadays, all these -omics data have to be integrated through bioinformatics tools and protocols for improving the discovery process from new hits to new use of drugs. In drug repositioning, Comparison of different -omics data from healthy and different pathological states could sometimes lead us to bring out defective or new pathways¹⁷. This allows us to elaborate some therapeutics scenarii with existing drug or combination or drugs targeting these pathways to restore the defective function^{5,8,10}.

Use of drug and disease gene expression data

Gene expression microarray data provide genome-wide expression levels of cell or tissues¹² of human diseases and has been widely used for classification of phenotypes⁹. It can permit to identify sets of genes that are up-regulated and down-regulated in a disease state compared to a normal state. These set is an expression *profile*, or *signature* of a disease. The large-scale comparison between disease signature and “disease treated by drugs” signature, allows discovering new connections between drugs and diseases by matching molecular profiles¹⁷ providing information for drug repositioning through target or pathway analysis. In their

study, *Sirota et al.*¹⁴ used a Significance Analysis of Microarrays (SAM) on GEO samples representing 100 tumor types from several human cancer cell lines treated with 164 different drugs to bring out drug-disease pathway relationship. From a statistical comparison of disease signatures to the reference drug expression signatures from the Connectivity Map, their work shown 16,000 possible drug-disease pairings and 2,664 were statistically significant ($q\text{-value} < 0.05$). Moreover, their methods provided significant candidate therapeutic drug repositioning for 53 out of the 100 diseases at a false discovery rate (FDR) of 5%. Their work permit to predict that HDAC inhibitors, classically used for the treatment of hematological tumor, could have therapeutic effects in the treatment of different types of brain tumors as well as other solid tumor (lung, colon, melanoma, etc.).

These omics approaches highlights the potential of gene transcription profiling to be the biological node to link chemistry, biology and the clinic¹¹ for drug repositioning⁴.

Disease phenotypes and secondary effects approaches

To predict drug repositioning, another approach is based on phenotypes and side-effects similarities between drugs and diseases. Side-effects of drugs are an important source of human phenotypic information. In this context, Khun et al.¹⁹ have developed an online public side-effect resource (SIDER) that connects 888 drugs to 1450 side effect terms and contains data on frequency in patients for one-third of the drug–side effect pairs.

Drug target identification using side-effect similarities

Drug repositioning is investigating the polypharmacology of drugs to identify new therapeutic uses for already approved drugs. Campillos et al.⁷elaborated a pipeline to link precise molecular data with their phenotypic observations. First, they established a classification between side-effect³ and drugs from clinical trial phases for identifying shared target proteins with drugs. This is based on the principle that drug leading to the same side effects might share molecular targets. Then they compiled a list of known drug-disease and their side-effects relationships to identify drugs which share the same side-effect profile but are prescribed within different therapeutic area. After creating a weighting scheme including a rareness weight and a correlation weight, the important result is that they observed an inverse correlation between

side-effect frequency and the likelihood of two drugs sharing a protein target. This result validated the side effects approach.

PREDICT

We already know that similar drugs are indicated for similar diseases PREDICT¹⁶ (Figure 3) is an algorithm for predicting drug indication based on the similarities between diseases by GBA (Guilt by association). If two diseases are treated by the same drug, another drug which treats only one disease might treat the other disease. For known associations, the algorithm of PREDICT construct drug-drug and disease-disease similarity based on data from multiple sources including the UMLS (Unified Medical Language System) database. The algorithm of PREDICT works in three phases: (i) from chemical structures, UMLS²⁸, Gene Ontology²⁹ and Human Proteome Project³⁰: building of drug–drug and disease–disease comparison set (ii) exploiting these sets of similarities to build classification features and a classification rule that distinguishes true from false drug–disease associations; and (iii) application of a logistic regression that automatically weighs the different features to get a classification score to predict new associations. In order to test whether predictions are in accordance with current experimental knowledge, there is a checking phase of the extent to which they appear in a clinical trial. The percentage of phase III clinical trial associations predicted by these methods is markedly high.

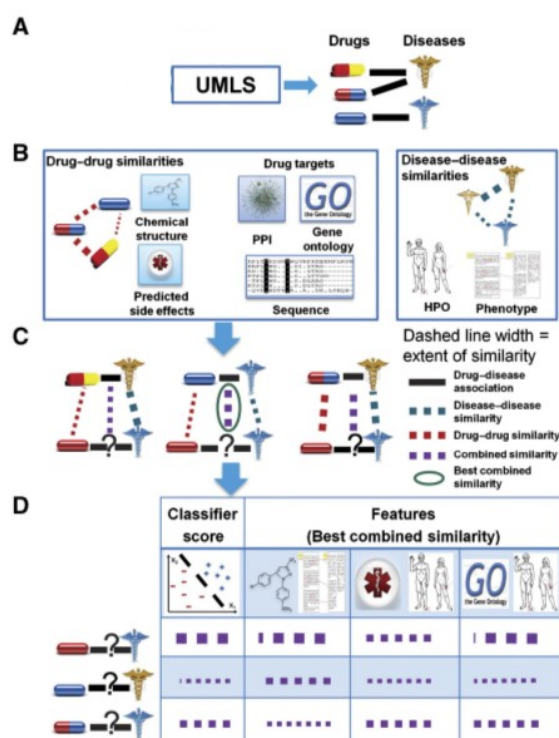


Figure 3. Pipeline of PREDICT algorithm. A. Formation of drug-disease association based on UMLS. B. Building of drug-drug and disease-disease similarity metrics. C. Evaluation of possible drug-drug association according to their similarities. D. Similarities integration into classification features and subsequent classification.

Concluding remarks

The pharmaceutical industry is under pressure to develop drugs in a context of crisis of new molecules lack. Drug repositioning is an interesting approach to get the most out of the safe compounds that already have been approved. First challenges of drug repositioning are the integration of large-scale data (structural knowledge on drugs/targets, -omics and phenotypic), which are very complex, and the computational capabilities to treat these richness of information. Depending on what type of data researchers have, each approach get own bias due to a lack of final clinical data. These biases are the next big challenge for developing new tools for these data integration and computation. Finally, all current approaches are applied to very limited set of data available for only hundreds of compound on the market especially at the phenotypic and clinical level. This would be beneficial if the large amount of clinical data accumulated by pharmaceutical industry for phase II and phase III and side-effects phase could be available in open data for drug repositioning research. We could understand fears of Big Pharma releasing these data about unexpected side-effects but it could be a big opportunity to rescue premature failure of compounds that show low efficacy in clinical trials. In the area of precision medicine, it is important to separate disease mechanism such as signaling network with or without treatment and off-target pathway to decipher the mechanism of action of drugs. Including the genetic polymorphisms specific to every patient, it could lead to the application of drug repositioning for personalized medicine. It would be greater to take in consideration this individual variability to scale drug repositioning from heterogeneity and complexity of patients while reducing inefficiency or toxicity of drugs.

Development of more powerful computational methods and pipelines are needed toward better drugs repositioning. This is the role of bioinformaticians to ensure high success rate of repositioned drugs.

References

1. Toplak et al. Assessment of Machine Learning Reliability Methods for Quantifying the Applicability Domain of QSAR

- Regression Models, journal of chemical information and modeling, 2014
2. Ye et al. Construction of drug network based on side effects and its application for drug repositioning. PLoS One 2014
 3. Duran-Frigola et al. P. Recycling side-effects into clinical markers for drug repositioning. Genome Med. 2012
 4. Von Eichborn, J. et al. PROMISCUOUS: a database for network-based drug-repositioning. Nucleic Acids Res. 2011
 5. Iorio, F., Isacchi, A., di Bernardo, D. & Brunetti-Pierri, N. Identification of small molecules enhancing autophagic function from drug network analysis. Autophagy 2010
 6. Keiser, M. J. et al. Predicting new molecular targets for known drugs. Nature 2009
 7. Campillos, M., Kuhn, M., Gavin, A.-C., Jensen, L. J. & Bork, P. Drug target identification using side-effect similarity. Science 2008
 8. Lussier, Y.A. and Chen, J.L. The emergence of genome-based drug repositioning. Sci. Transl. Med. 2011
 9. Nevins, J.R. and Potti, A. Mining gene expression profiles: expression signatures as cancer phenotypes. Nat. Rev. Genet. 2007
 10. Hieronymus, H. et al. (2006) Gene expression signature-based chemical genomic prediction identifies a novel class of HSP90 pathway modulators. Cancer Cell 10, 321–330
 11. Lamb, J. et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science 2006
 12. Dudley JT, Tibshirani R, Deshpande T, Butte AJ. Disease signatures are robust across tissues and experiments. Mol Syst Biol. 2009
 13. M.J. Keiser, B.L. Roth, B.N. Armbruster, P. Ernsberger, J.J. Irwin, B.K. Shoichet, Nat. Biotechnol. 25, 2007
 14. Sirota et al. Discovery and preclinical validation of drug indications using compendia of public gene expression data Sci Transl Med. 2011
 15. Eming et al. Drug Target Prediction and Repositioning Using an Integrated Network-Based Approach, Plos One 2013
 16. Goettlieb et al. PREDICT: a method for inferring novel drug indications with application to personalized medicine, Mol Sys Bio, 2011
 17. Hu et al. Human Disease-Drug Network Based on Genomic Expression Profiles, PLoS One. 2009
 18. <http://www.drugrepositioningconference.com/index/index.php/agenda/2014-agenda>
 19. Khun et al. A side effect resource to capture phenotypic effects of drugs, Mol Syst Biol 2010
 20. <http://www.pharmext.com/>
 21. <http://www.biovista.com/>
 22. <http://www.gene.com/>
 23. <http://www.meliordiscovery.com/>
 24. <http://www.aurorabiomed.com/>
 25. Gfeller et al SwissTargetPrediction: a web server for target prediction of bioactive small molecules. Nucleic acid research, 2014
 26. <http://www.ncbi.nlm.nih.gov/geo/>
 27. <https://www.broadinstitute.org/cmap/>
 28. <http://www.nlm.nih.gov/research/umls/>
 29. <http://geneontology.org/>
 30. <http://www.hupo.org/human-proteome-project/>

Acknowledgments

Thanks to Anaïs Baudot for her help and feedback on the manuscript.

Tab1. Example of existing drugs successfully repositioned

Drug	Original indication	New indication
Amantadine	Influenza	Parkinson's disease
Amphotericin	Fungal Infections	Leishmaniasis
Aspirin	Inflammation, Pain	Antiplatelet
Atomoxetine	Antidepressant	ADHD
Bromocriptine	parkinson's disease	Diabetes Mellitus
Bupropion	Depression	Smoking cessation
Colchicine	Gout	Recurrent pericarditis
Gabapentin	Epilepsy	Neuropathic pain
Methotrexate	Cancer	Psoriasis, Rheumatoid arthritis
Miltefosine	Cancer	Visceral Leishmaniasis
Propanolol	Hypertension	Migraine prophylaxis
Retinoic acid	Acne	Acute promyelocytic leukemia
Ropinirole	Parkinson's disease	Restless leg syndrome
Zidovudine	Cancer	HIV/AIDS